Identifying Demand Effects in a Large Network of Product Categories

Sarah Gelper

Innovation, Technology Entrepreneurship & Marketing, Eindhoven University of Technology

Ines Wilms

Faculty of Economics and Business, Katholieke Universiteit Leuven

Christophe Croux

Faculty of Economics and Business, Katholieke Universiteit Leuven

Abstract. Planning marketing mix strategies requires retailers to understand within- as well as cross-category demand effects. Most retailers carry products in a large variety of categories, leading to a high number of such demand effects to be estimated. At the same time, we do not expect cross-category effects between all categories. This paper presents sparse estimation for cross-category effects in a large network of categories using the Vector Autoregressive Market Response Model. Sparse estimation increases the interpretability of the estimation results and leads to more accurate estimation and prediction. We show that cross-category effects are much more prevalent than previous literature suggests. We show that there are important differences in the responsiveness of demand and in the influence of marketing mix variables between categories and identify perceived complements and substitutes.

Keywords. Cross-category demand effects; Market response model; Sparse estimation; Vector autoregressive model

1 Introduction

While within-category demand effects of the marketing mix have been studied extensively, cross-category effects are less well understood (Leeflang and Selva, 2012). Cross-category effects might be substantial for a number of reasons. Some categories are complements, e.g. bacon and eggs as studied by Niraj et al. (2008) or cake mix and cake frosting studied by Manchanda et al. (1999), while others are substitutes, e.g. frozen, refrigerated and shelf-stable juices (Wedel and Zhang, 2004). But cross-effects also exist among categories that are not complements or substitutes. First, price promotions in one category alleviate the budget constraint such that consumers are able to spend more on other, possibly unrelated, categories (Song and Chintagunta, 2007; Lee et al., 2013). Second, advertising and promotions generate more store traffic and therefore more sales in other categories (Bell et al., 1998). And third, as a result of brand extensions, brands are no longer limited to one category (Kamkura and Kang, 2007; Ma et al., 2012). So advertising and promotion of a brand within one category might spill over to own brand sales in other categories.

While cross-category effects might be substantial for these reasons, we do not expect that each category's marketing mix variables influence each and every other category. Instead, we expect some cross-category effects to be zero – or very close to zero – but we don't know which. In order to identify which cross-category effects are substantial, this paper presents sparse estimation of the vector autoregressive (VAR) model. The estimation is *sparse* in the sense that some parameters in the model can be estimated as exactly zero. This allows us to easily identify categories that influence one another, also when many categories are considered together.

Initiated by the work of Baghestani (1991) and Dekimpe and Hanssens (1995), the VAR Market Response Model has become a standard, flexible tool to measure own- and cross-effects of marketing actions in a competitive environment. The main drawback of the VAR model is the risk of overparametrization because the number of parameters increases quadratically with the number of included categories. Previous studies on cross-category effects have therefore limited their attention to a small number of categories by studying substitutes or complements (Kamkura and Kang, 2007; Song and Chintagunta, 2007; Leeflang et al., 2008; Bandyopadhyay, 2009; Ma et al., 2012). We present an estimation technique for cross-category effects in large product category networks with many parameters and even few data points. Short observation periods are commonplace in marketing practice since many firms discard data that are older than one year (Lodish and Mela, 2007).

The remainder of this article is organized as follows. Section 2 presents the methodology. We describe in detail the sparse estimation of the VAR model. We then discuss how to construct impulse response functions. We compare the sparse estimation technique with two Bayesian estimators in the same Section 2. In Section 3, a simulation study shows the excellent performance of the proposed methodology in terms of estimation reliability and prediction accuracy. Section 4 presents our data and model. Section 5 present our findings on cross-category demand effects. We first identify which categories are most responsive to changes in other categories and which are most influential, and then identify perceived substitutes and complements based on estimated cross-price elasticities. Finally, Section 6 concludes.

2 Methodology

Identifying cross-category demand effects using standard methods is challenging because the sheer number of such effects makes them hard to estimate. To overcome an explosion of the number of parameters to be estimated in the VAR model, marketing researchers have used pre-estimation dimension reduction techniques, i.e. they first impose restrictions on the model and then estimate the reduced model. Four such common pre-estimation dimension reduction techniques are (i) treating marketing variables as exogenous (e.g. Nijs et al., 2001; Pauwels et al., 2002 and Nijs et al., 2007), (ii) estimating submodels rather than a full model (e.g. Srinivasan et al., 2000; Srinivasan et al., 2004), (iii) aggregating or pooling over, for instance, stores or competitors (e.g. Horvath et al., 2005; Slotegraaf and Pauwels, 2008), and (iv) applying Least Squares to a restricted model (e.g. Dekimpe and Hanssens, 1995;

Dekimpe et al., 1999; Nijs et al., 2007). Most researchers applying pre-estimation dimension reduction techniques recognize that they do so because of the practical limitations of standard estimation techniques rather than for theoretical reasons (e.g. Srinivasan et al., 2004 and Bandyopadhyay, 2009).

In situations where the number of parameters to estimate is large relative to the sample size, the Lasso proposed by Tibshirani (1996) provides a solution within the multiple regression model. The Lasso minimizes the least squares criterion penalized for the sum of the absolute values of the regression parameters. This penalization forces some of the estimated regression coefficients to be exactly zero, which results in selection of the pertinent variables in the model. The Lasso method is well established (Buhlmann and van de Geer, 2011; Chatterjee and Lahiri, 2011) and shows good performance in various applied fields (Wu et al., 2009; Fan et al., 2011).

The Lasso technique can not be directly applied to the VAR model because the VAR model differs from a multiple regression model in two important aspects. First, a VAR model contains several equations, corresponding to a multivariate regression model. Correlations between the error terms of the different equations need to be taken into account. Second, a VAR model is dynamic, containing lagged versions of the same time series as right-hand side variables of the regression equation. The sparse estimator used in this paper extends the lasso method such that it can deal with both these aspects of the VAR. It builds further on a sparse estimator of the multivariate regression model (Rothman et al., 2010), and the groupwise lasso for categorical variables (Yuan and Lin, 2006; Meier et al., 2008).

2.1 Model Specification

Sales, price and marketing activity are measured for several categories over a certain time period. We collect all these time series in a multivariate time series $\mathbf{y}_t$ with q components. In our empirical study on cross-category demand effects, $\mathbf{y}_t$ contains sales, price and marketing activity for 17 product categories, hence $q = 3 \times 17 = 51$. The VAR Market Response Model is given by

$$\mathbf{y}_t = B_1\mathbf{y}_{t-1} + B_2\mathbf{y}_{t-2} + \ldots + B_p\mathbf{y}_{t-p} + \mathbf{e}_t\,, \tag{1}$$

where p is the lag length. The autoregressive parameters B_1 to B_p are $(q \times q)$ matrices, which capture both within- and cross-category effects. The elements of these matrices measure the effect of sales, price and marketing activity of one category on the sales, price and marketing activity of other categories (including its own). The error term $\mathbf{e}_t$ is assumed to follow a $N_q(0, \Sigma)$ distribution. We assume, without loss of generality, that all time series are mean centered such that no intercept is included.

If the number of components q in the multivariate time series is large, the number of unknown elements in the sequence of matrices $B_1, \ldots, B_p$ explodes to pq^2, and accurate estimation by standard methods is no longer possible. Sparse estimation, with many elements of the matrices $B_1, \ldots, B_p$ estimated as zero, brings an outcome: it will not only provide estimates with smaller mean squared error, but also substantially improve model interpretability. The method we propose does not require the researcher to prespecify which entries in the B_j matrices are zero and which are not. Instead, the estimation and variable selection are simultaneously performed. This is particularly of interest in situations where there is no a priori information on which time series is driving which.

The instantaneous correlations in model (1) are captured in the error covariance matrix Σ. If the dimension q is large relative to the number of observations, estimation of Σ becomes problematic. The estimated covariance matrix risks getting singular, i.e. its inverse does not exist. Hence, we also induce sparsity in the estimation of the inverse error covariance matrix $\Omega = \Sigma^{-1}$. The elements of Ω have a natural interpretation as partial correlations between the error components of the q equations in model (1). If the ij-th element of the inverse covariance matrix is zero this means that, conditional on the other error terms, there is no correlation between the error terms of equations i and j.

2.2 Penalized Likelihood Estimation

This section defines the sparse estimation procedure for the VAR model. The Sparse VAR estimator is defined by minimizing a measure of goodness-of-fit to the data combined with a *penalty* for the magnitude of the model parameters. It is convenient to first recast model (1) in stacked form as

$$y = X\beta + e\,, \tag{2}$$

where y is a vector of length nq containing the stacked values of the time series. If the multivariate time series has length T, then $n = T - p$ is the number of time points for which all current and lagged observations are available. The vector β contains the stacked vectorized matrices $B_1, \ldots, B_p$, and e the vector of stacked error terms. The matrix $X = I_q \otimes X_0$, with $X_0 = (\underline{\mathbf{Y}}_1, \ldots, \underline{\mathbf{Y}}_p)$, is of dimension $(nq \times pq^2)$. Here $\underline{\mathbf{Y}}_j$ is an $(n \times q)$ matrix, containing the values of the q series at lag j in its columns, for $1 \leq j \leq p$, with p the maximum lag. The symbol $\otimes$ stands for the Kronecker product.

The sparse estimator of the autoregressive parameters β and the inverse covariance matrix $\Omega = \Sigma^{-1}$ are obtained by minimizing the negative log likelihood with a groupwise penalization on the β and a penalization on the off-diagonal elements of Ω:

$$(\hat{\beta}, \hat{\Omega}) = \underset{(\beta,\Omega)}{\operatorname{argmin}} \frac{1}{n}(y - X\beta)'\tilde{\Omega}(y - X\beta) - \log|\Omega| + \lambda_1 \sum_{g=1}^{G} ||\beta_g||_2 + \lambda_2 \sum_{k \neq k'} |\Omega_{kk'}|\,, \tag{3}$$

where $||u||_2 = (\sum_{i=1}^n u_i^2)^{1/2}$ is the Euclidean norm and $\tilde{\Omega} = \Omega \otimes I_n$. The vector β_g in (3) is a subvector of β, containing the regression coefficients for the lagged values of the same time series in one of the q equations in model (1). The coefficients of the lagged values of the same time series form a group. The total number of groups is $G = q^2$ because there are q groups within each of the q equations. The penalty on the regression coefficients enforces that either all elements of the group $\hat{\beta}_g$ are zero or none. As a result, the estimated B_j matrices, for $j = 1, \ldots, p$, have their zero elements in exactly the same cells. The penalization on the off-diagonal elements of Ω induces sparsity in the estimate $\hat{\Omega}$. Finally, the scalars λ_1 and λ_2 control the degree of sparsity of the regression estimator and the inverse covariance matrix estimator, respectively. The larger these values, the more sparsity is imposed.

The estimator is consistent for the unknown model parameters, see Meier et al. (2008) and Friedman et al. (2008). Our approach is similar to Hsu et al. (2008) who use the Lasso within a VAR context. However, they do not account for the group-structure in the VAR model, nor do they impose sparsity on the error covariance matrix. Details on the algorithm to perform penalized likelihood estimation and the selection of the sparsity parameters λ_1 and λ_2 can be found in Appendix A.

2.3 Bayesian Estimators

An alternative to the sparse estimation technique is to impose prior information in a Bayesian setting. Bayesian regularization techniques have been proposed for the VAR model in Litterman (1980) and are used in various applied fields such as macroeconomics (Gefang, 2014; Banbura et al., 2010), finance (Carriero et al., 2012) and marketing (Lenk and Orme, 2009; Horvath and Fok, 2013; Bandyopadhyay, 2009). They are also applicable to a situation like ours where there are many parameters to be estimated with a limited observation period, and are thus a good benchmark. However, these methods are not sparse, they do not perform variable selection simultaneously with model estimation. The following two paragraphs elaborate on two Bayesian estimators which serve as non-sparse alternatives.

Minnesota Prior. The original Minnesota prior only specifies a prior distribution for the regression parameters of the VAR model. The error covariance matrix Σ is assumed to be diagonal, and estimated by $\hat{\Sigma}_{ii} = \hat{\sigma}_i^2$ with $\hat{\sigma}_i^2$ the standard OLS estimate of the error variance in an AR(p) model for the i^{th} time series (Koop and Korobilis, 2009). The prior distribution of the regression parameters is taken to be multivariate normal:

$$\beta \sim N(\underline{\beta}_M, \underline{V}_M). \tag{4}$$

For the prior mean, the common choice is $\underline{\beta}_M = 0_{Kq}$ for stationary series. The prior covariance matrix $\underline{V}_M$ is diagonal. The posterior distribution is again multivariate normal. Full technical details can be found in Koop and Korobilis (2009).

The main advantage of the Minnesota prior is its ease of implementation, since posterior inference only involves the multivariate normal distribution. However, imposing the Minnesota prior only ensures that the parameter estimates are *shrunken* towards zero, while the Sparse VAR ensures that some parameters will be estimated as *exactly* zero.

Normal Inverted Wishart Prior. The Minnesota prior takes the error covariance matrix Σ as fixed and diagonal and, hence, not as an unknown parameter. To overcome this problem, Banbura et al. (2010) impose an inverse Wishart prior on the Σ matrix. More precisely,

$$\beta \mid \Sigma \sim N(\underline{\beta}_{NIW}, \Sigma \otimes \Omega_0) \quad \text{and} \quad \Sigma \sim iW(S_0, \nu_0), \tag{5}$$

where $\underline{\beta}_{NIW}, \Omega_0, S_0$ and ν_0 are hyperparameters. Under this normal inverted Wishart prior (labeled in the remainder of this paper as "NIW"), the posterior for β, conditional on Σ is normal, and the posterior for Σ is again inverted Wishart. Full technical details can be found in Banbura et al. (2010).

2.4 Impulse Response Functions

Impulse response functions (IRFs) are extensively used to assess the dynamic effect of external shocks to the system such as changes in the marketing mix. An IRF pictures how a change to a certain variable at moment t impacts the value of any other time series at time $t+k$, accounting for interrelations with all other variables. The magnitude of the effect is plotted as a function of k. An extensive discussion on the interpretation of the IRF in marketing modeling can be found in Dekimpe and Hanssens (1995). We use IRFs to gain insight in the dynamics of within and cross-category sales, promotion and price effects on each of the 17 product category sales. The IRFs are easily computed as a function of the Sparse VAR estimator (see Hamilton, 1991). Since we want to account for correlated error terms, we use generalized IRFs (Pesaran and Shin, 1998; Dekimpe and Hanssens, 1999).

To obtain confidence bounds for the generalized IRFs estimated by Sparse VAR, we use a residual parametric bootstrap procedure (Chatterjee and Lahiri, 2011). We generate

$N_b = 1000$ time series of length T from the VAR model (2). The invertible estimate of Σ delivered by the Sparse VAR estimation procedure is needed to draw random numbers for the $N_q(0, \Sigma)$ error distribution. For each of these N_b multiple time series, the estimates of the regression parameters are computed. We compute the covariance matrix of the N_b bootstrap replicates. For each of the N_b generated series impulse response functions are computed; the 90% confidence bounds are then obtained by taking the 5% and 95% percentiles.

3 Estimation and prediction performance

We conduct a simulation study to compare the proposed Sparse VAR with Bayesian methods using the Minnesota and NIW prior. As benchmarks, we include the classical Least Squares (LS) estimator and two restricted versions of LS which are often used in practice. In the 1-step Restricted LS (Dekimpe and Hanssens, 1995; Dekimpe et al., 1999), we estimate the model with classical LS, delete all variables with $|t\text{-statistic}| \leq 1$, and re-estimate the model with the remaining variables. We also consider an iterative Restricted LS method described in Lutkepohl and Kratzig (2004) where we fit the full model using LS and sequentially eliminate the variables leading to the largest reduction of BIC until no further improvement is possible, of which a close variant was used by Nijs et al. (2007).

We simulate from a VAR model with $q = 10$ dimensions and $p = 2$ lags. Each time series has an own auto-regressive structure and we include system dynamics among the different series. The first series leads series two to five, while the sixth series leads time series 7 to 10. Specifically, the data generating processes are given by

$$\mathbf{y}_t = \begin{bmatrix} B_1 & 0 \\ 0 & B_1 \end{bmatrix} \mathbf{y}_{t-1} + \begin{bmatrix} B_2 & 0 \\ 0 & B_2 \end{bmatrix} \mathbf{y}_{t-2} + \mathbf{e}_t \,,$$

with

$$B_1 = \begin{bmatrix} 0.4 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.4 & 0.4 & 0.0 & 0.0 & 0.0 \\ 0.4 & 0.0 & 0.4 & 0.0 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.4 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.0 & 0.4 \end{bmatrix} \text{ and } \quad B_2 = \begin{bmatrix} 0.2 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.2 & 0.2 & 0.0 & 0.0 & 0.0 \\ 0.2 & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.2 & 0.0 & 0.0 & 0.2 & 0.0 \\ 0.2 & 0.0 & 0.0 & 0.0 & 0.2 \end{bmatrix}.$$

In total, there are $pq^2 = 200$ regression parameters to be estimated with 36 true parameter values different from zero. The 10-dimensional error term $\mathbf{e}_t$ is drawn from a multivariate normal with mean zero and covariance matrix $\Sigma = 0.1 I_{10}$. We generate $N_s = 1000$ multivariate time series of length 50 according to the above simulation scheme.

Performance measures We evaluate the different estimators in terms of (i) estimation accuracy, (ii) sparsity recognition performance, and (iii) forecast performance.

To evaluate estimation accuracy, we compute the mean absolute estimation error (MAEE), averaged over the simulation runs and over the 200 parameters

$$\text{MAEE} = \frac{1}{N_s} \frac{1}{pq^2} \sum_{s=1}^{N_s} \sum_{j=1}^{p} \sum_{k,l=1}^{q} |\hat{b}_{klj}^s - b_{klj}|,$$

where $\hat{b}_{klj}^s$ is the estimate of b_{klj}, the kl^{th} element of the matrix B_j corresponding to lag j, for the s^{th} simulation run.

Concerning sparsity recognition, we compute the true positive rate and true negative rate

$$\text{TPR}(\hat{b}, b) = \frac{\#\{(k,l,j) : \hat{b}_{klj} \neq 0 \textit{ and } b_{klj} \neq 0\}}{\#\{(k,l,j) : \ b_{klj} \neq 0\}}$$

$$\text{TNR}(\hat{b}, b) = \frac{\#\{(k,l,j) : \hat{b}_{klj} = 0 \textit{ and } b_{klj} = 0\}}{\#\{(k,l,j) : \ b_{klj} = 0\}}.$$

The true positive rate (TPR) gives an indication on the number of true relevant regression parameters detected by the estimation procedure. The true negative rate (TNR) measures the hit rate of detecting a true zero regression parameter. Both should be as large as possible.

Finally, we conduct an out-of-sample rolling window forecasting exercise. Using the same simulation design as before, we generate multivariate time series of length $T = 60$, and use

Table 1: Mean Absolute Estimation Error (MAEE), True Positive Rate (TPR), True Negative Rate (TNR) and Mean Absolute Forecast Error (MAFE), averaged over 1000 simulation runs, are reported for every method.

Method	MAEE	TPR	TNR	MAFE
Sparse VAR	0.041	0.860	0.848	0.359
LS	0.157	1	0	0.540
Restricted LS: 1-step	0.121	0.709	0.541	0.520
Restricted LS: Iterative	0.116	0.261	0.775	0.516
Bayseian: Minnesota	0.044	1	0	0.355
Bayseian: NIW	0.077	1	0	0.476

a rolling window of length $S = 50$. For all estimation methods, 1-step-ahead forecasts are computed for $t = S, \ldots, T-1$. Next, we compute the Mean Absolute Forecast Error (MAFE), averaged over all time series and across time

$$\text{MAFE} = \frac{1}{T-S}\frac{1}{q}\sum_{t=S}^{T-1}\sum_{i=1}^{q} \mid \hat{y}_{t+1}^{(i)} - y_{t+1}^{(i)} \mid, \tag{6}$$

where $y_{t+1}^{(i)}$ is the value of the i^{th} time series at time $t+1$.

Results Table 1 presents the performance measures of the Sparse VAR, the Bayesian and benchmark methods. The Sparse VAR estimator performs best in terms of estimation accuracy. It attains the lowest value of the MAEE (0.041). A paired t-test confirms that the Sparse VAR significantly outperforms the other methods (all p-values < 0.001).

It might seem surprising that we can beat the LS in terms of estimation accuracy, since the LS estimator is the Maximum Likelihood estimator, hence asymptotically efficient, and the best linear unbiased estimator (BLUE). However, the Sparse VAR is not a linear estimator and is meant to be used when a limited amount of data are available for estimating a complex model, in particular for shorter time series. If the length of the time series would tend to infinity, then the LS estimator will finally end up having the smallest MAEE.

Sparsity recognition performance is evaluated using the true positive rate and the true negative rate, reported in Table 1. For the LS and Bayesian estimators, all parameters are estimated as non-zero, resulting in a perfect true positive rate and zero true negative rate. Among the variable selection methods, the Sparse VAR performs best. Sparse VAR achieves a value of the true positive rate of 0.86; 0.85 for the true negative rate.

Finally, we evaluate the forecast performance of the different estimators by the Mean Absolute Forecast Error in Table 1. The Sparse VAR and the Bayesian estimator with Minnesota prior achieve the best forecast performance. A Diebold-Mariano test confirms that these two methods perform significantly better than the others (p-values < 0.001).

4 Data and Model

We use the sparse estimation technique for large VARs described in Section 2 to identify cross-category demand effects across 17 categories in the Dominick's Finer Foods database. This database is a well-established source of weekly scanner data from a large Midwestern supermarket chain, Dominick's Finer Foods (see e.g. Kamkura and Kang, 2007; Pauwels, 2007). We first describe the data and model in more detail, and then report on the insights the Sparse VAR generates.

We use all 17 product categories in the Dominick's Finer Foods database containing food and drink items, a much broader selection of categories than previous studies on cross-category demand effects have considered. A description of each product category can be found in Table 2. We analyze data on the product category level since this is most relevant for retailers to study the impact of promotion and price effects (Ailawadi et al., 2009; Leeflang and Selva, 2012). For 15 stores, we obtain weekly sales, pricing and promotional feature and display data for the 17 product categories.

Sales. Category sales volumes for the 17 categories, measured in dollars per week.

Promotion. The promotional data include the percentage of SKUs of each category that are promoted (feature and display) in a given week, following Srinivasan et al. (2004).

Prices. To aggregate pricing data from the SKU level to the product category level, we follow Srinivasan et al. (2004) and Pauwels et al. (2002) in using SKU market shares as weights. Prices are not deflated because there is strong evidence that people are sensitive to

Table 2: Description of the 17 categories from Dominick's Finer Foods database that are analyzed in this paper.

Category	Description	Category	Description	Category	Description
BER	Beer	SNA	Snack Crackers	CER	Cereals
BJC	Bottled Juices	CIG	Cigarettes	OAT	Oatmeal
RFJ	Refrigerated Juices	FEC	Front-end-candies	FRD	Frozen Dinners
FRJ	Frozen Juices	COO	Cookies	FRE	Frozen Entrees
SDR	Soft Drinks	CHE	Cheeses	TNA	Canned Tuna
CRA	Crackers	CSO	Canned Soup		

Table 3: Description of the 15 data sets. Each data set contains multivariate time series for sales ($\mathbf{Y}_t$), promotion ($\mathbf{M}_t$) and prices ($\mathbf{P}_t$).

Store	Number of	Dimension			
	Time Points	$\mathbf{Y}_t$	$\mathbf{M}_t$	$\mathbf{P}_t$	Total
Store 1-15	77	17	16	17	50

nominal rather than real price changes (Shafir et al., 1997) over short time periods.

We use data from January 1993 to July 1994, 77 weeks in total. We neither use observations after 1994 since Srinivasan et al. (2004) pointed out that manufacturers made extensive use of 'pay-for-performance' price promotions as of 1994, which are not fully reflected in the Dominick's database, nor data before 1993 since they contain missing observations. This data range is short relative to the dimension of the VAR, which calls for a regularization approach such as the Sparse VAR. For all stores, we collect data on sales, promotion and pricing for all 17 categories. Only for cigarettes, no promotion variable is included in the VAR since none of the SKUs in that category were promoted during the observation period.

We estimate a separate VAR model for each store, which allows to evaluate the robustness of the findings. The multivariate time series entering the VAR model are the log-differenced sales ($\mathbf{Y_t}$), differenced promotion ($\mathbf{M_t}$), and log-differenced prices ($\mathbf{P_t}$).[1] The dimensions of the time series are represented in Table 3. We use the Vector Autoregressive model, with

[1]Following standard practice, we first test for stationarity. A stationarity test of all individual time series using the Augmented Dickey-Fuller test indicates that most time series in levels are integrated of order 1.

endogenous promotion and prices,

$$\begin{bmatrix} \mathbf{Y_t} \\ \mathbf{P_t} \\ \mathbf{M_t} \end{bmatrix} = B_0 + B_1 \begin{bmatrix} \mathbf{Y_{t-1}} \\ \mathbf{P_{t-1}} \\ \mathbf{M_{t-1}} \end{bmatrix} + \ldots + B_p \begin{bmatrix} \mathbf{Y_{t-p}} \\ \mathbf{P_{t-p}} \\ \mathbf{M_{t-p}} \end{bmatrix} + \mathbf{e_t}. \tag{7}$$

Averaged across stores, the selected value of p is 2 for the Sparse VAR. Also for the Bayesian estimators, the lag order of the VAR is selected using the BIC criterion, which is 1 for the majority of the stores.

5 Empirical Results

We first discuss the findings on the direct effects that occur in the product category network as measured by the estimated regression parameters. A direct sales effect between category A and category B occurs when category's A sales, price or promotion immediately influence the sales of category B. Then we turn to the complete chain of direct and indirect effects using Impulse Response Functions. Indirect effects involve other variables in the system. For instance, an indirect sales effect of category A on category B occurs when the sales, promotion or prices of category A influence (the sales, promotion or prices in) a certain other category C which, in turn, influences the sales of category B. Since we work in a time series setting, both direct and indirect effects might be dynamic in the sense that the effect can occur with a certain delay.

5.1 A network of product categories

We analyze cross-category demand effects as a network of interlinked product categories. Recently, network perspectives have been increasingly used by marketing researchers to model, for example, the network value of a product in a product network (Oestreicher-Singer et al., 2013) or to investigate the flow of influence in a social network (Zubcsek and Sarvary, 2011). In our case, the 17 product categories are the nodes of the network. If the Sparse VAR estimation results indicate, by giving a non-zero estimate, that sales in

Table 4: Proportion of nonzero within and cross-category sales, promotion and price effects, averaged across 15 stores and 17 product categories.

	Sales	Promotion	Price
Within-category	96%	30%	34%
Cross-category	21%	21%	19%

one category have a direct influence on sales in another, a directed edge is drawn between them. The resulting network is plotted in Figure 1. Similarly, Figures 2 and 3 present cross-promotion and cross-price effects. If promotion or price in one category directly influence sales in another category, respectively, this is indicated by a directed edge.

First note that, as expected, the cross-category networks are sparse – not each category influences each and every other category. While the sparse VAR estimation favors zero-effects, it does not enforce them. Here, as many as 78% of all estimated effects are zero-effects. Table 4 summarizes the prevalence of within-and cross-category effects. As expected, within-category effects are more common than cross-category effects. For all categories, past values of the own category's sales are selected for almost all stores. Cross-category effects of sales on sales (21%), promotion on sales (21%) and price on sales (19%) are about equally prevalent.

Next, we focus on category responsiveness and influence. Category responsiveness is measured by the number of edges pointing towards a category. It helps us understand which categories are responsive to changes in other categories. Category influence, on the other hand, is measured by the number of edges originating from a category. Category influence identifies which categories are important drivers of other category's sales.

The most responsive categories to changes in sales in other categories are oatmeal (OAT) and crackers (CRA), each having four incoming edges in Figure 1. Frozen dinners (FRD) is the most responsive to promotion (cfr. three incoming edges in Figure 2) while bottled juices (BJC) and canned soup (CSO) are the most responsive to price in other categories (cfr. respectively four and five incoming edges in Figure 3). Especially the oatmeal (OAT)

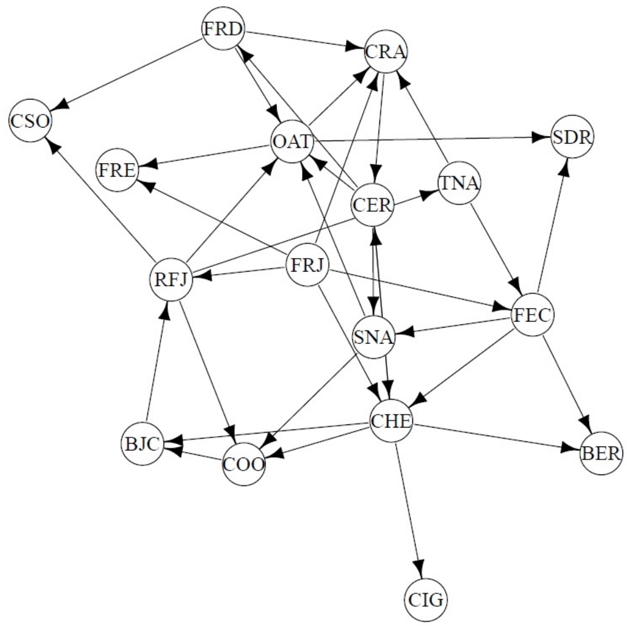

Figure 1: Cross-category sales effects: a directed edge is drawn from one category to another if its sales influences sales in the other category for the majority of stores.

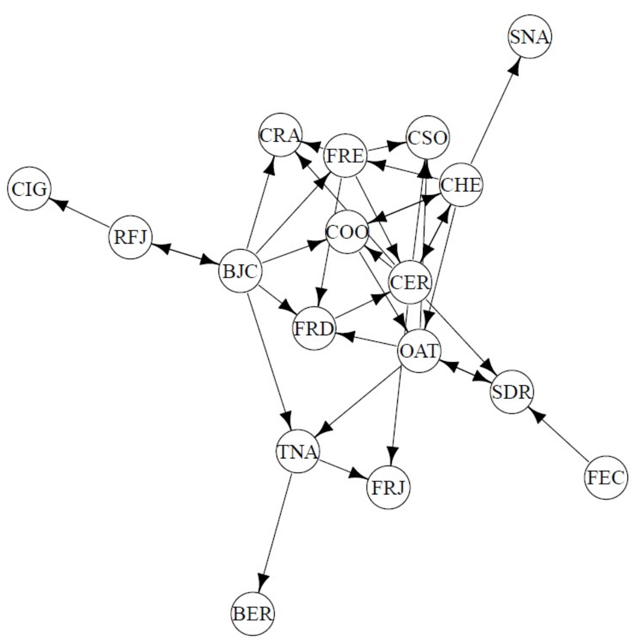

Figure 2: Cross-category promotion effects: a directed edge is drawn from one category to another if its promotion influences sales in the other category for the majority of stores.

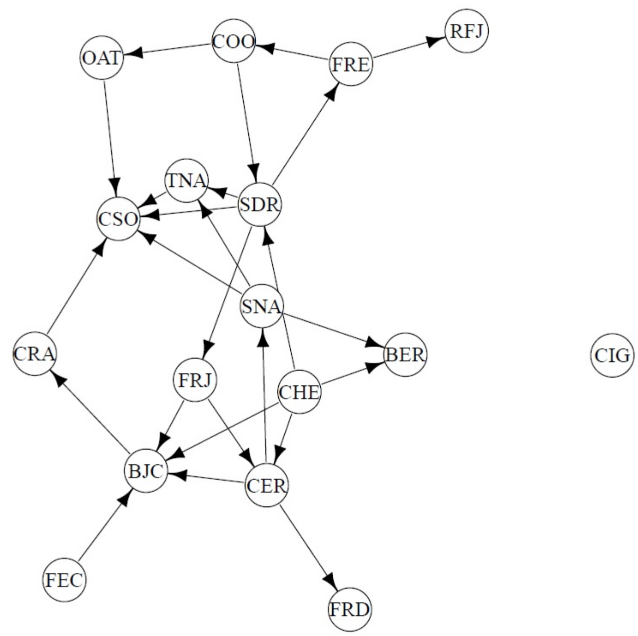

Figure 3: Cross-category price effects: a directed edge is drawn from one category to another if its price influences sales in the other category for the majority of stores.

category is highly responsive to changes in other categories; its sales are influenced by sales in four other categories (Figure 1), promotion in three other categories (Figure 2) and price in one other category (Figure 3). The cigarettes (CIG) category on the other hand is the least responsive (cfr. one, one and zero incoming edges in Figures 1 , 2 and 3 respectively). The finding that cigarettes are a non-responsive category is not surprising as this is in line with cigarettes being addictive, hence, people who smoke will buy cigarettes no matter what.

Concerning category influence, changes in sales of cheeses (CHE) influences sales in five other categories on average (cfr. five outgoing edges in Figure 1). In terms of the marketing mix, changes in promotion of cereals (CRE) and in the prices of cheeses (CHE) influence the highest number of sales in other categories (cfr. respectively six and five outgoing edges in Figure 2 and 3). Furthermore, changes in the cigarettes category do not influence any other category.

To confirm the robustness of the results obtained by Sparse VAR, we check whether the sales, promotion and price results on category responsiveness and influence are consistent

Table 5: Kendall's coefficient of concordance across stores of sales, promotion and price cross-category effects on sales for both category responsiveness and influence. P-values are indicated between parentheses.

	Sales	Promotion	Price
Responsiveness	0.17 (<0.001)	0.16 (0.001)	0.30 (<0.001)
Influence	0.30 (<0.001)	0.56 (<0.001)	0.40 (<0.001)

Table 6: Sizes of within and cross-category effects of sales, promotion and price on sales, summed across 10 lags of IRFs, averaged across stores and in absolute value.

	Sales	Promotion	Price
Within-category	0.057	0.006	0.004
Cross-category	0.002	0.005	0.002

across stores. We compute Kendall's coefficient of concordance W on the graphs in Figures 2-4 at the store level. As W increases from 0 to 1, there is stronger consistency across stores. Table 5 indicates that all values of Kendall's W are significant, indicating the categories that are responsive to changes in other categories and those that are important drivers of other category's sales are consistent across stores.

5.2 Impulse Response Functions

We compute effect sizes from IRFs by summing the responses across the first 10 lags of the IRF. In Table 6, we report the within and cross-category sales, promotion and price effect sizes, averaged across stores. We report absolute effect sizes in order not to average out positive and negative effects. Table 6 indicates that, for example, a one standard deviation shock in prices leads to an accumulated absolute change of .004 in own sales growth over a time period of 10 lags. As expected, we systematically find that within-category effects are larger in magnitude than cross-category effects, especially so for sales and prices. For the marketing mix, promotions exert stronger within- as well as cross-category effects than price changes.

Second, the IRFs give estimated directions to investigate the source of cross-category ef-

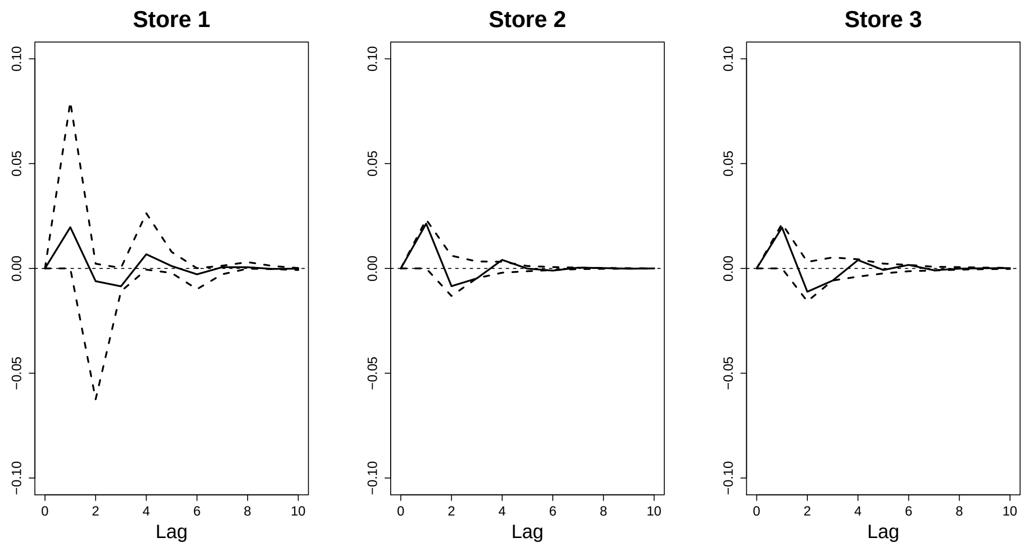

Figure 4: Impulse response function: response of frozen juices sales growth to a one standard deviation impulse in the price of soft drinks.

fects, in particular whether consumers perceive categories as complements or as substitutes. Following the standard economic definition (Pashigian, 1998), complements are defined as goods having a negative cross-price elasticity, whereas substitutes are defined as goods having a positive cross-price elasticity. For each store, we calculate the cross-price elasticity between all 17 product categories using the Sparse VAR. Table 7 reports the five main perceived complements and substitutes ranked according to the sizes of the estimated cross-price elasticities. The most prominent cross-category price effect is observed for soft drinks. An increase in soft drink prices makes consumers spend more on other drinks, in particular frozen juices, and spend less on other categories, in particular on frozen entrees. This indicates that the budgetary constraint is an important source of cross-category effects because soft drinks account for a relatively large proportion of the total retail expenditures of US families (22% of retail spending in our data) – price changes in such an important category tighten the budget constraint.

The joint dynamic effect of a one standard deviation price increase of soft drinks on the sales growth of frozen juices is depicted in Figure 4 for the first three stores in the data

Table 7: Main perceived complements and substitutes: total effect size summed across 10 lags of IRFs and averaged across stores for which the effect is selected, and number of stores for which the effect is selected (out of 15).

Impulse in price of	Response of sales of	Number of stores	Effect
Complements			
Soft Drinks	Frozen Entrees	14	-0.0196
Canned Tuna	Canned Soup	14	-0.0187
Cereals	Frozen Dinners	13	-0.0121
Cheeses	Beer	11	-0.0100
Bottled Juices	Crackers	8	-0.0128
Substitutes			
Front-end-candies	Bottled Juices	12	0.0150
Cookies	Oatmeal	12	0.0075
Soft Drinks	Frozen Juices	11	0.0085
Snack Crackers	Beer	14	0.0063
Frozen Juices	Bottled Juices	8	0.0046

Table 8: Mean Absolute Forecast Error (MAFE) for category-specific sales, averaged over the 15 stores, the 17 product categories, and across time. P-values of a Diebold-Mariano test comparing the Sparse VAR to its alternatives are indicated between parentheses.

	Sparse VAR		Restricted LS		Bayesian Methods	
		LS	1-step	Iterative	Minnesota	NIW
MAFE	736.80	1298.54 (<0.01)	784.96 (<0.01)	734.82 (0.38)	875.47 (<0.01)	1078.03 (<0.01)

set. We see a sharp increase in frozen juices sales growth one week after the soft drink price increase, indicating substitution. However, the next two weeks, sales growth of frozen juices slows down, which could indicate stockpiling behavior (Gangwar et al., 2014).

5.3 Forecast Performance

Although prediction is not the main goal of the proposed methodology, we deem it important to show that the Sparse VAR can compete with other methods in terms of prediction accuracy. We estimate model (7) for each store and perform a forecast exercise (cfr. Section 3), using a rolling window of length $S = 67$. 1-step-ahead forecasts of sales for each product category are computed for $t = S, \ldots, T - 1$, with $T = 77$. The same estimation methods as in Section 3 are used.

Results on the sales predictions are summarized in Table 8 by the Mean Absolute Forecast Error (MAFE), averaged across time and over the 17 product categories and 15 stores. The variable selection methods Sparse VAR, 1-step Restricted LS and Iterative Restricted LS perform, on average, better than the methods that don't perform variable selection. This indicates that sparsity improves prediction accuracy. The Sparse VAR and Iterative Restricted LS estimator achieve the best forecasting performance. A Diebold-Mariano test confirms that these two methods significantly outperform the other methods. We conclude that the improvement in interpretability of the model obtained by Sparse VAR, as discussed in the previous section, does not come at the cost of lower forecast performance.

6 Discussion

This paper presents a Sparse VAR methodology to detect the inter-relationships in a large product category network. We show that an important number of cross-category demand effects are detected for a large number of categories. We identify perceived complements and substitutes based on estimated price elasticities and conclude that cross-category effects also exist between categories that are not directly related at first sight. This illustrates the relevance and need to study – potentially a large number of – product categories simultaneously. Identification of cross-category effects is important for both marketing mix planning and inventory management to prevent potentially harmful reactions of consumers to stock-outs (Fitzsimons, 2000; Campo et al., 2004). Even though most categories affect at least one other category, on average they affect only two. This indicates that, while cross-category effects are prevalent, many of them are absent, calling for a sparse estimation procedure that succeeds in highlighting the main inter-relationships in the product category network.

The advantage of the Sparse VAR is that it overcomes the dimensionality problem. We show that this leads to more accurate estimation and prediction results as compared to standard Least Squares methods. An alternative to Sparse VAR is Bayesian regularization by imposing priors on the parameters. The benefit of Sparse VAR over the Bayesian methods is the improvement in terms of interpretability of the results due to the sparseness. If the researcher wishes to restrict some of the parameters to zero a priori, using marketing theory, this is of course still possible to implement with the Sparse VAR. The same holds for the reverse, i.e. forcing some variables to be included in the model, which can be done by adjusting the penalty on the regression coefficients in (3).

The methodology presented in this paper is relevant in a variety of other settings. First, the Sparse VAR can be used to study competitive demand effects across many competitors. As was shown, the VAR model is ideal for measuring competitive effects because it is able to capture own- and cross-elasticity of sales to both pricing and marketing spending (Srinivasan et al., 2004; Horvath et al., 2005). However, typically only three competitors are included in such studies, while using the Sparse VAR allows for a larger number to be included. Second, in the field of international marketing research there is an increased interest in studying cross-country spill-over effects, as for example in Albuquerque et al. (2007), van Everdingen et al. (2009) and Kumar and Krishnan (2002). Every country that is added to the data set leads to an increase in the number of cross-country parameters to be estimated. Using the proposed methodology, a large VAR model could be built which allows spill-over effects between many countries. Finally, the Market Response Model could be extend with data on online word of mouth or online search, which are now readily available. Especially in the Big Data era, most companies have an abundance of data (Chintagunta et al., 2013), such that large VAR models will become even larger as more granular data becomes available. Estimation of complex models with limited amounts of data opens a rich area for future research.

Acknowledgments. Financial support from the FWO (Research Foundation Flanders) is gratefully acknowledged (FWO, contract number 11N9913N).

Appendix A Penalized Likelihood Estimation

We iteratively solve the minimization problem (3) for β conditional on Ω and then for Ω conditional on β.

Solving for $\beta|\Omega$: When Ω is fixed, the minimization problem in (3) is equivalent to minimizing

$$\hat{\beta}|\Omega = \underset{\beta}{\operatorname{argmin}} \frac{1}{n}(\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta) + \lambda_1 \sum_{g=1}^{G} ||\beta_g||_2 \,, \tag{A.1}$$

where $\tilde{y} = Py$, $\tilde{X} = PX$, and P is a matrix such that $P'P = \tilde{\Omega}$. The transformation of the data to $\tilde{y}$ and $\tilde{X}$ ensures that the resulting model has uncorrelated and homoscedastic error terms. The above minimization problem is equivalent to the groupwise lasso of Yuan and Lin (2006), implemented in the R package `grplasso` (Meier, 2009).

Solving for $\Omega|\beta$: When β is fixed, the minimization problem in (3) reduces to

$$\hat{\Omega}|\beta = \underset{\Omega}{\operatorname{argmin}} \frac{1}{n}(y - X\beta)'\tilde{\Omega}(y - X\beta) - \log|\Omega| + \lambda_2 \sum_{k \neq k'} |\Omega_{kk'}| \,, \tag{A.2}$$

which corresponds to penalized covariance estimation. Using the glasso algorithm of Friedman et al. (2008), available in the R package `glasso` (Friedman et al., 2011), the optimization problem in (A.2) is solved.

We start the algorithm by taking $\widehat{\Omega} = I_q$ and iterate until convergence. We iterate until $max_s|\hat{\beta}_{s,i} - \hat{\beta}_{s,i-1}| < \epsilon$, with $\hat{\beta}_{s,i}$ the s^{th} parameter estimate in iteration i (same for $\hat{\Omega}$) and the tolerance ϵ set to 10^{-3}.

Selecting the Sparsity Parameters and the order of the VAR We first determine the optimal values of λ_1 and λ_2 for a fixed value of p, the order of the VAR. The sparsity parameters λ_1 and λ_2 are selected according to a minimal Bayes Information Criterion (BIC). In the iteration step where β is estimated conditional on Ω, we solve (A.1) over a range of values for λ_1 and select the one with lowest value of

$$BIC_{\lambda_1} = -2 \log L_{\lambda_1} + k_{\lambda_1} \log(n), \tag{A.3}$$

where L_{λ_1} is the estimated likelihood, corresponding to the first term in (A.1), using sparsity parameter λ_1. Furthermore, k_{λ_1} is the number of non-zero estimated regression coefficients and n the number of observations. Similarly, for selecting λ_2, we use the BIC given by

$$BIC_{\lambda_2} = -2 \log L_{\lambda_2} + k_{\lambda_2} \log(n)\,. \tag{A.4}$$

Finally, the order p of the VAR is selected such that the lowest value of the BIC, with sparsity parameters λ_1 and λ_2 set at their optimal values, is attained.

References

Ailawadi, K.; Beauchamp, J.; Donthu, N.; Gauri, D. and Shankar, V. (2009), "Communication and promotion decisions in retailing: A review and directions for future research," *Journal of Retailing*, 85(1), 42–55.

Albuquerque, P.; Bronnenberg, B. and Corbett, C. (2007), "A spatio-temporal analysis of global diffusion of ISO 9000 and ISO 14000 certification," *Management Science*, 53, 451–468.

Baghestani, H. (1991), "Cointegration analysis of the advertising-sales relationship," *Journal of Industrial Economics*, 39, 671–681.

Banbura, M.; Giannone, D. and Reichlin, L. (2010), "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92.

Bandyopadhyay (2009), "A dynamic model of cross-category competition: Theory, tests and applications," *Journal of Retailing*, 85, 468–479.

Bell, D.; Ho, T. and Tang, C. (1998), "Determining where to shop: Fixed and variable costs of shopping," *Journal of Marketing Research*, 35, 352–369.

Buhlmann, P. and van de Geer, S. (2011), *Statistics for high-dimensional data: methods, theory and applications*, Springer Series in Statistics, ISBN 978-3-642-20192-9.

Campo, K.; Gijsbrechts, E. and Nisol, P. (2004), "Dynamics in consumer response to product unavailability: do stock-out reactions signal response to permanent assortment reductions?" *Journal of Business Research*, 57, 834–843.

Carriero, A.; Kapetanios, G. and Marcellino, M. (2012), "Forecasting government bond yields with large Bayesian vector autoregressions," *Journal of Banking & Finance*, 36, 2026–2047.

Chatterjee, A. and Lahiri, S. (2011), "Bootstrapping lasso estimators," *Journal of the American Statistical Association*, 106, 608–625.

Chintagunta, P.; Hanssens, D. and Hauser, J. (2013), "Call for Papers. Marketing Science special issue on big data: Integrating marketing, statistics, and computer science," *Marketing Science*, 32, 678.

Dekimpe, M. and Hanssens, D. (1995), "The persistence of marketing effects on sales," *Marketing Science*, 14, 1–21.

— (1999), "Sustained spending and persistent response: A new look at long-term marketing profitability," *Journal of Marketing Research*, 36, 397–412.

Dekimpe, M.; Hanssens, D. and Silva-Risso, J. (1999), "Long-run effects of price promotions in scanner markets," *Journal of Econometrics*, 89, 269–291.

Fan, J.; Lv, J. and Qi, L. (2011), "Sparse high-dimensional models in economics," *Annual Review of Economics*, 3, 291–317.

Fitzsimons, G. (2000), "Consumer Response to Stockouts," *Journal of Consumer Research*, 27, 249–266.

Friedman, J.; Hastie, T. and Tibshirani, R. (2008), "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, 9, 432–441.

— (2011), *glasso: Graphical lasso- estimation of Gaussian graphical models*, r package version 1.4.

Gangwar, M.; Kumar, N. and Rao, R. (2014), "Consumer stockpiling and competitive promotional strategies," *Marketing Science*, 33, 94–113.

Gefang, D. (2014), "Bayesian double adaptive elastic-net Lasso for VAR shrinkage," *International Journal of Forecasting*, 30, 1–11.

Hamilton, J. (1991), *Time Series Analysis*, Princeton University Press.

Horvath, C. and Fok, D. (2013), "Moderating factors of immediate, gross, and net cross-brand effects of price promotions," *Marketing Science*, 32, 127–152.

Horvath, C.; Leeflang, P.; Wieringa, J. and Wittink, D. (2005), "Competitive reaction- and feedback effects based on VARX models of pooled store data," *International Journal of Research in Marketing*, 22, 415–426.

Hsu, N.-J.; Hung, H.-L. and Chang, Y.-M. (2008), "Subset selection for vector autoregressive processes using Lasso," *Computational Statistics and Data Analysis*, 52, 3645–3657.

Kamkura, W. and Kang, W. (2007), "Chain-wide and store-level analysis for cross-category management," *Journal of Retailing*, 83, 159–170.

Koop, G. and Korobilis, D. (2009), "Bayesian multivariate time series methods for empirical macroeconomics," *Foundations and Trends in Econometrics*, 3, 267–358.

Kumar, V. and Krishnan, T. (2002), "Multinational diffusion models: An alternative approach," *Marketing Science*, 21, 318–330.

Lee, S.; Kim, J. and Allenby, M. (2013), "A direct utility model for asymmetric complements," *Marketing Science*, 32, 454–470.

Leeflang, P. and Selva, J. (2012), "Cross-demand effects of price promotions," *Journal of Academy of Marketing Science*, 40, 572–586.

Leeflang, P.; Selva, J.; Van Dijk, A. and Wittink, D. (2008), "Decomposing the sales promotion bump accounting for cross-category effects," *International Journal of Research in Marketing*, 25, 201–214.

Lenk, P. and Orme, B. (2009), "The value of informative priors in Bayesian inference with sparse data," *Journal of Marketing Research*, 46, 832–845.

Litterman, R. (1980), "A Bayesian procedure for forecasting with vector autoregression," *Working Paper, Massachusetts Institute of Technology, Dept. of Economics.*

Lodish, L. and Mela, C. (2007), "If brands are built over years, why are they managed over quarters?" *Harvard Business Review*, 85, 158.

Lutkepohl, H. and Kratzig, M. (2004), *Applied time series econometrics*, Cambridge University Press.

Ma, Y.; Seetharaman, P. and Narasimhan, C. (2012), "Modeling Dependencies in brand choice outcomes across complementary categories," *Journal of Retailing*, 88, 47–62.

Manchanda, P.; Ansari, A. and Gupta, S. (1999), "The shopping basket: A model for multicategory purchase incidence decisions," *Marketing Science*, 18, 95–114.

Meier, L. (2009), *grplasso: Fitting user specified models with group lasso penalty*, r package version 0.4-2.

Meier, L.; van de Geer, S. and Buhlmann, P. (2008), "The group lasso for logistic regression," *Journal of the Royal Statistical Society, Series B*, 70, 53–71.

Nijs, V.; Dekimpe, M.; Steenkamp, J. and Hanssens, D. (2001), "The category demand effects of price promotions," *Marketing Science*, 20, 1–22.

Nijs, V.; Srinivasan, S. and Pauwels, K. (2007), "Retail-price drivers and retailer profits," *Marketing Science*, 26, 473–487.

Niraj, R.; Padmanabhan, V. and Seetharaman, P. (2008), "A cross-category model of households' incidence and quantity decisions," *Marketing Science*, 27, 225–235.

Oestreicher-Singer, G.; Libai, B.; Sivan, L.; Carmi, E. and Yassin, O. (2013), "The network value of products," *Journal of Marketing*, 77, 1–14.

Pashigian, B. (1998), *Price theory and applications*, McGraw-Hill Higher Education.

Pauwels, K. (2007), "How retailer and competitor decisions drive the long-term effectiveness of manufacturer promotions for fast moving consumer goods," *Journal of Retailing*, 83, 297–308.

Pauwels, K.; Hanssens, D. and Siddarth, S. (2002), "The long-term effects of price promotions on category incidence, brand choice and purchase quantity," *Journal of Marketing Research*, 39, 421–439.

Pesaran, H. and Shin, S. (1998), "Generalized impulse response analysis in linear multivariate models," *Economics Letters*, 58, 17–29.

Rothman, A.; Levina, E. and Zhu, J. (2010), “Sparse multivariate regression with covariance estimation,” *Journal of Computational and Graphical Statistics*, 19, 947–962.

Shafir, E.; Diamond, P. and Tversky, A. (1997), “Money illusion,” *Quarterly Journal of Economics*, 112, 341–374.

Slotegraaf, R. and Pauwels, K. (2008), “The impact of brand equity and innovation on the long-term effectiveness of promotions,” *Journal of Marketing Research*, 45, 293–306.

Song, I. and Chintagunta, P. (2007), “A discrete-continuous model for multicategory purchase behaviour of households,” *Journal of Marketing Research*, 44, 595–612.

Srinivasan, S.; Pauwels, K.; Hanssens, D. and Dekimpe, M. (2004), “Do promotions benefit manufacturers, retailers, or both?” *Management Science*, 50, 617–629.

Srinivasan, S.; Popkowski Leszczyc, P. and Bass, F. (2000), “Market share response and competitive interaction: The impact of temporary, evolving and structural changes in prices,” *International Journal of Research in Marketing*, 17, 281–305.

Tibshirani, R. (1996), “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Series B*, 58, 267–288.

van Everdingen, Y.; Fok, D. and Stremersch, S. (2009), “Modeling global spillover of new product takeoff,” *Journal of Marketing Research*, 46, 637–652.

Wedel, M. and Zhang, J. (2004), “Analyzing brand competition across subcategories,” *Marketing Science*, 41, 448–456.

Wu, T.; Chen, Y. and Hastie, T. (2009), “Genome-wide association analysis by lasso penalized logistic regression,” *Bioinformatics*, 25, 714–721.

Yuan, M. and Lin, Y. (2006), “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society Series B*, 68, 49–67.

Zubcsek, P. and Sarvary, M. (2011), "Advertising to a social network," *Quantitative Marketing and Economics*, 9, 71–107.